

## Comment to the Editor

### On Imposing Detailed Balance in Complex Reaction Mechanisms

In a recent article, Colquhoun et al. (Colquhoun, D., K. A. Dowsland, M. Beato, and A. J. Plested. 2004. *Biophys. J.* 86:3510–3518) present a method to impose detailed balance on a complex reaction mechanism, which relies on identifying reaction cycles. We propose a method to impose detailed balance without explicitly considering reaction cycles. In our method, the detailed balance constraint is imposed by parameterizing the model in terms of the concentrations and balanced fluxes at equilibrium. A general rate matrix satisfying detailed balance is derived from these parameters. We illustrate this technique for voltage- and ligand-dependent single molecule kinetics. In the single molecule case, when ligand binding obeys the law of mass action, we point out that our parameterization correctly gives the ligand dependence of the equilibrium probabilities without solving any equations. We also show how to impose detailed balance for general nonlinear mass-action kinetics. The techniques for obtaining a minimal reaction network subject to the detailed balance constraint are also presented and illustrated on a large network.

#### INTRODUCTION

The principle of microscopic reversibility implies detailed balance—the statement that, at thermodynamic equilibrium, each individual reaction is balanced. That is, at equilibrium each individual reaction occurs with equal forward and backward fluxes. A reaction system that satisfies detailed balance does not consume or dissipate free energy at thermodynamic equilibrium (1). Although reaction systems with no closed loops (or acyclic systems) always satisfy detailed balance, most complex reaction schemes involve reaction cycles, and satisfying detailed balance requires that the product of equilibrium constants around a reaction cycle equals one. Colquhoun et al. (2) recently presented methods to impose detailed balance on complex reaction mechanisms, such as ion channel kinetic schemes modeled with finite Markov chains. The methods rely on finding a fundamental cycle basis with respect to a spanning tree of a graph representing the reaction topology. In this comment, we discuss an alternative method that balances the forward and backward fluxes of each reversible reaction and does not require direct consideration of reaction cycles. The method will be discussed for single molecule dynamics and mass-action

kinetics. We will also discuss techniques for identifying a minimal reaction network for imposing detailed balance.

#### SINGLE MOLECULE DYNAMICS

Idealized single molecules, such as an ion channel with a finite number of states, are often modeled by homogeneous Markov processes, which can in general be described by the so-called forward equation,

$$\frac{dP(t)}{dt} = P(t)Q,$$

where element  $p_{ij}(t)$  in matrix  $P(t)$  is the probability that the system is in state  $j$  at time  $t$ , provided it was in state  $i$  at time 0, and  $Q$  is called the generator matrix, whose element  $q_{ij}$  denotes the rate constant for the transition from state  $i$  to  $j$ ,  $i \neq j$ . Each diagonal entry  $q_{ii} = -\sum_{j \neq i} q_{ij}$ , so that  $Qu = 0$ , where  $u$  is a column vector containing all ones. Detailed balance holds at equilibrium implying that all fluxes balance:  $w_i q_{ij} = w_j q_{ji}$  (3,4), where  $w$  is the equilibrium distribution of states such that  $w^T Q = 0$ . Defining a diagonal matrix  $W = \text{diag}(w)$ , the detailed balance conditions are written as

$$WQ = (WQ)^T. \quad (1)$$

Imposing detailed balance on a reaction network requires that  $Q$  satisfies Eq. 1. It is evident that the matrix  $Q_s = WQ$  is a symmetric generator matrix and entries in  $Q_s$  are equilibrium fluxes.  $Q$  can be written in terms of  $Q_s$  and  $W$  as follows:

$$Q = W^{-1}Q_s. \quad (2)$$

Note that  $W$  need not be normalized for  $Q$  to obey detailed balance. Normalization can be incorporated in  $Q_s$ . However, if  $W$  is normalized then it gives the equilibrium distribution and  $Q_s$  gives the equilibrium fluxes. By specifying the equilibrium distribution (entries in  $W$ ) and the equilibrium fluxes (entries in  $Q_s$ ), one can impose detailed balance on  $Q$  using the above equation, which in fact provides an alternative parameterization scheme to that of directly specifying a system using rate constants in  $Q$ . This way of specifying a system is especially useful in situations where equilibrium distribution and fluxes can be obtained from measured data. Constraints in addition to detailed balance, such as measured rate constants or dependence between rate constants, which can be expressed by equations that are linear with respect to the logarithms of rate constants, are

Submitted August 3, 2005, and accepted for publication April 24, 2006.

Address reprint requests to John E. Pearson, Tel.: 505-667-7585; E-mail: pearson@lanl.gov.

© 2006 by the Biophysical Society

0006-3495/06/08/1136/06 \$2.00

doi: 10.1529/biophysj.105.071852

also linear with respect to the logarithms of the components of  $Q_s$  and  $W$ . Therefore, the technique to calculate constrained parameters in terms of free parameters using linear equations proposed by Colquhoun et al. (2) and Qin et al. (5) also applies.

In practice, a model reaction scheme is given a priori. The topology of this scheme can be encoded within  $Q_s$  (i.e., a nonzero entry indicates a reaction and a zero entry indicates no reaction), and topology is preserved in Eq. 2. In the case of finding a  $Q$  matrix that best fits data, via a maximum likelihood method (6), Eq. 2 can be substituted directly into the likelihood formula. The following outlines a procedure for using Eq. 2 in a maximum likelihood method (assuming no other constraints):

1. Specify a symmetric generator  $Q_s$  with arbitrary positive entries consistent with a given reaction scheme. Specify an arbitrary equilibrium distribution vector  $w$ .
2. Calculate  $Q = W^{-1}Q_s$ , where  $W = \text{diag}(w)$ .
3. For fitting data using a maximum likelihood method, maximize the likelihood function by adjusting  $w$  and fluxes in  $Q_s$  iteratively, until satisfactory parameters are found.

The values  $Q_s$  and  $w$  can be enforced to be positive during the course of model fitting by using constrained optimization or following Qin et al. (5).

*Parameter dependence: ligand concentration, voltage*

The above program can be carried out in the case that there is dependence on other physical parameters such as ligand concentration, pressure, temperature, voltage, etc. In principle, one can choose  $Q_s$  and  $W$  to be any desired functions of the parameters.

Note that if detailed balance is to be satisfied for all physical parameters, the ratio of the product of rate constants one way around any loop to the product the other way around the loop must necessarily be independent of parameter values. Here we give a physically motivated prescription for enforcing detailed balance, without considering loops, for reactions with exponential voltage dependence and ligand dependence governed by the law of mass action. Mass action reactions have monomial ligand dependence, so in the case of a single ligand with concentration  $[L]$ ,

$$q_{ij} = \kappa_{ij}[L]^{\alpha_{ij}} e^{\beta_{ij}V},$$

where  $\alpha_{ij}$  are integers, and  $\kappa_{ij}$  and  $\beta_{ij}$  are constants. This formulation can be made to enforce positivity of  $q_{ij}$  (which is sometimes not desirable (7)) by replacing  $\kappa_{ij}$  with  $e^{\mu_{ij}}$  (5). From Eq. 2, we know that under detailed balance, each rate constant  $q_{ij}$  is the ratio of the equilibrium flux  $q_{ij}^s$  to the equilibrium probability  $w_i$ . Parameterizing  $q_{ij}^s$  as  $\kappa_{ij}^s [L]^{\alpha_{ij}^s} e^{\beta_{ij}^s V}$  ( $i \neq j$ ,  $q_{ii}^s = -\sum_{j \neq i} q_{ij}^s$ ) and  $w_i$  as  $k_i [L]^{a_i} e^{b_i V}$  results in the factorization

$$q_{ij} = \frac{q_{ij}^s}{w_i} = \frac{\kappa_{ij}^s}{k_i} \frac{[L]^{\alpha_{ij}^s}}{[L]^{a_i}} \frac{e^{\beta_{ij}^s V}}{e^{b_i V}}, \tag{3}$$

where  $\kappa_{ij}^s$ ,  $\beta_{ij}^s$ , and  $\alpha_{ij}^s$  are symmetric in  $i$  and  $j$ . This equation provides a formula for imposing detailed balance in a model with ligand- and voltage-dependence. If positivity is desired,  $\kappa_{ij}^s$  and  $k_i$  can be written as exponentials (5):  $\kappa_{ij}^s = \exp(\mu_{ij}^s)$  and  $k_i = e^{\delta_i}$ , so that  $\ln(q_{ij}) = \mu_{ij}^s - \delta_i + (\alpha_{ij}^s - a_i) \ln(L) + (\beta_{ij}^s - b_i)V$ . For mass-action kinetics,  $a_i$  is the number of ligands bound in the  $i^{\text{th}}$  state and  $\alpha_{ij}^s$  is the total number of ligands present in the reversible reaction from state  $i$  to state  $j$ . (To be specific, if the state  $i$  has  $p$  ligands bound and state  $j$  has  $p + r$  ligands bound and the reaction is  $S_i + nL \rightleftharpoons S_j + (n-r)L$ , then  $\alpha_{ij}^s = p + n$  for mass action.) The  $b_i$  and  $\beta_{ij}$  are the net charge displacements along the electric field in state  $i$  and the transition state between states  $i$  and  $j$ , respectively. With this formulation, the  $w_i$  values are specified in terms of explicit parameters of the model, thus giving the equilibrium probabilities, up to a normalization scalar that is introduced below. The case of more than one species of ligand is handled similarly and treated in the following example.

We illustrate how to impose detailed balance on the de Young-Keizer model (8) of the IP3 receptor, which is a calcium channel that binds two ligands,  $\text{Ca}^{2+}$  and IP3. The model, which has six fundamental cycles, is shown in Fig. 1.

The equilibrium vector with proper ligand dependence is simply written down according to the prescription,

$$w_i = k_i C^{a_{iC}} I^{a_{iI}},$$

where the ligand concentrations  $[\text{Ca}^{2+}]$  and  $[\text{IP3}]$  are denoted as  $C$  and  $I$ , respectively, and where  $a_{iC}$  and  $a_{iI}$  are the number of calcium ions and IP3 molecules, respectively,

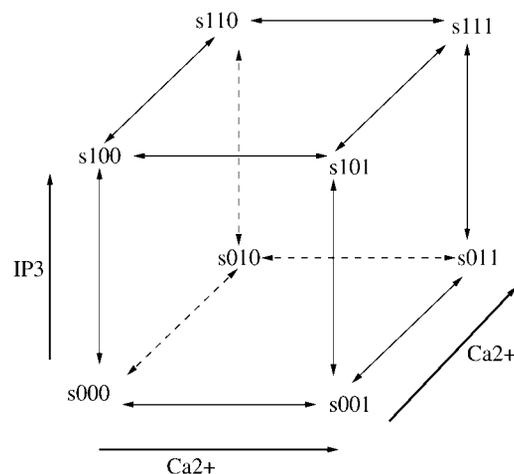


FIGURE 1 State transition model of an IP3 receptor in the endoplasmic reticulum with calcium dependence. The ion channel is gated by two ligands: calcium ( $[\text{Ca}^{2+}]$ ) and inositol 1,4,5-trisphosphate (IP3).

bound in state  $i$  ( $i = 0, \dots, 7$ ). For the de Young-Keizer model,

$$w = \frac{1}{\eta} [1, k_1C, k_2C, k_3C^2, k_4I, k_5IC, k_6IC, k_7IC^2],$$

where  $\eta$  is the normalization scalar defined by  $\sum_i w_i = 1$ :

$$\eta = 1 + k_1C + k_2C + k_3C^2 + k_4I + k_5IC + k_6IC + k_7IC^2.$$

Note that we have arrived at the formula for  $w_i$ , including  $w_6 = k_6IC$ , which exhibits the celebrated bell-shaped calcium-dependence of the open probability in the de Young-Keizer model (8), without actually solving any equations.

For mass-action kinetics involving noncatalytic ligands, the exponents  $\alpha_{ij}^s$ , which give the ligand dependence of  $Q_s$ , obey

$$\alpha_{ij}^s = \max(a_i, a_j). \quad (4)$$

Symmetry of  $\alpha^s$  follows from the symmetry of the max function. For models with multiple ligands, Eq. 4 holds for each ligand separately. For the de Young-Keizer model,

$$\alpha_{ijI}^s = \max(a_{iI}, a_{jI})$$

$$\alpha_{ijC}^s = \max(a_{iC}, a_{jC}).$$

The fluxes are then given by

$$Q_s = \frac{1}{\eta} \begin{pmatrix} x_0 & \kappa_{01}C & \kappa_{02}C & 0 & \kappa_{04}I & 0 & 0 & 0 \\ \kappa_{01}C & x_1 & 0 & \kappa_{13}C^2 & 0 & \kappa_{15}IC & 0 & 0 \\ \kappa_{02}C & 0 & x_2 & \kappa_{23}C^2 & 0 & 0 & \kappa_{26}IC & 0 \\ 0 & \kappa_{13}C^2 & \kappa_{23}C^2 & x_3 & 0 & 0 & 0 & \kappa_{37}IC^2 \\ \kappa_{04}I & 0 & 0 & 0 & x_4 & \kappa_{45}IC & \kappa_{46}IC & 0 \\ 0 & \kappa_{15}IC & 0 & 0 & \kappa_{45}IC & x_5 & 0 & \kappa_{57}IC^2 \\ 0 & 0 & \kappa_{26}IC & 0 & \kappa_{46}IC & 0 & x_6 & \kappa_{67}IC^2 \\ 0 & 0 & 0 & \kappa_{37}IC^2 & 0 & \kappa_{57}IC^2 & \kappa_{67}IC^2 & x_7 \end{pmatrix},$$

where the diagonal entries  $x_i$  are minus the sum of the other elements in each row, as is required for a generator.

## DEGREES OF FREEDOM

To calculate the number of free parameters in Eq. 2, let  $N_s$  and  $N_r$  denote the numbers of states and reversible reactions (i.e., the number of links in the state-transition diagram),

respectively. Without any constraint, the transition matrix  $Q$  contains  $2N_r$  free parameters. If detailed balance is the only physical constraint that the system must obey, the number of free parameters is  $N_r + N_s - 1$  because there are  $N_r$  free parameters in the symmetric matrix  $Q_s$  and  $N_s - 1$  in  $w$  (with a normalization constraint  $w^T u = 1$ ). Colquhoun et al. (2) obtained the same number of free parameters by counting the number of connections constrained by detailed balance around a set of fundamental cycles ((2), their Eq. 9). Detailed balance still holds for  $Q$  even if  $w$  is not normalized because a multiplication constant that normalizes  $w$  can always be factored out of  $Q_s$ . Here we consider the example shown in Colquhoun et al. ((2), their Fig. 3). For that scheme,  $Q$  can be written as

$$Q = W^{-1}Q_s$$

$$= W^{-1} \begin{pmatrix} d_{11} & x_1 & 0 & x_2 & 0 & x_3 & 0 & 0 \\ x_1 & d_{22} & x_4 & 0 & 0 & 0 & x_5 & 0 \\ 0 & x_4 & d_{33} & x_6 & 0 & 0 & 0 & x_7 \\ x_2 & 0 & x_6 & d_{44} & x_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_8 & d_{55} & x_9 & 0 & x_{10} \\ x_3 & 0 & 0 & 0 & x_9 & d_{66} & x_{11} & 0 \\ 0 & x_5 & 0 & 0 & 0 & x_{11} & d_{77} & x_{12} \\ 0 & 0 & x_7 & 0 & x_{10} & 0 & x_{12} & d_{88} \end{pmatrix},$$

where each  $x_i$  denotes a nonzero entry. The diagonal entries  $d_{ii}$  assure that every row sums to zero. The parameters are the  $x_i$  and the  $w_i$ . In agreement with Colquhoun et al. (2),  $W$  and  $Q_s$  have 19 free parameters: the 12  $x_i$  and 7 independent  $w_i$ .

## MASS-ACTION KINETICS

Here we consider a general biochemical network with mass-action rate laws. An idealized reaction model is implied, i.e., all reactions proceed in an isothermal and well-mixed container of a constant volume. Whereas before we treated ligand concentrations as parameters, now the concentration of each participant is a dynamic variable. A network has a set of  $m$  biochemical species (e.g., molecules, proteins and protein

complexes, etc.)  $\mathbf{x} = \{x_i | i = 1, \dots, m\}$ . A reaction-group  $G$  is defined as a set of species participating in a reaction, either as reactants or products. Any reaction within the network is a process that converts reactants into products, and a unidirectional reaction is denoted as a group transition:  $G_r \rightarrow G_p$ . Note that reactants in one reaction could be products in another. Assuming the network has  $n$  distinct reaction groups, we can compile information about stoichiometry into the  $m \times n$  stoichiometric matrix  $S$ . The components  $s_{ij}$  are the stoichiometric coefficients for species  $i$  in reaction group  $j$  (note that  $s_{ij} \geq 0$ ). All the reaction groups can be listed using an  $n$  by 1 column vector  $\mathbf{g}(\mathbf{x})$ , where each entry  $g_i(\mathbf{x}) = \prod_{j=1}^n x_j^{s_{ij}}$ . A biochemical network is characterized by kinetic rate constants for the reactions. This kinetic information can be compiled into an  $n \times n$  kinetics matrix  $K$ , where  $k_{ij} (i \neq j)$  is the kinetic rate constant for the reaction that converts reaction group  $i$  into group  $j$ . For a mass-action rate law,  $g_i(\mathbf{x})k_{ij}$  is the corresponding reaction rate. The diagonal entry  $k_{ii}$  of  $K$  consolidates rate constants for all the outgoing fluxes from reaction group  $g_i(\mathbf{x})$ , i.e.,  $k_{ii} = -\sum_{j,j \neq i} k_{ij}$ . The flux rates for all the groups can be expressed as a vector  $\mathbf{f} = K^T \mathbf{g}(\mathbf{x})$  and the whole system can be described by the following dynamic equation,

$$\frac{d\mathbf{x}}{dt} = S\mathbf{f} = SK^T \mathbf{g}(\mathbf{x}). \quad (5)$$

This description of general mass-action kinetics was developed by Horn and Jackson (9). Notations adopted here are similar to those of Chaves et al. (10). The detailed balance constraint can be described as  $g_i(\mathbf{x}_{\text{eq}})k_{ij} = g_j(\mathbf{x}_{\text{eq}})k_{ji}$ , where  $\mathbf{x}_{\text{eq}}$  denotes the equilibrium concentrations. The constraint can also be written in matrix form as

$$GK = (GK)^T,$$

where  $G = \text{diag}(\mathbf{g}(\mathbf{x}_{\text{eq}}))$ . Thus,

$$K = G^{-1}K_s, \quad (6)$$

where  $K_s = GK$  is a symmetric matrix. To impose detailed balance on this system using the method described above, one can first specify the equilibrium concentration vector  $\mathbf{x}_{\text{eq}}$ , subsequently construct vector  $\mathbf{g}(\mathbf{x}_{\text{eq}})$ , and then use Eq. 6 to calculate  $K$ . Analogous to the analysis for single molecule dynamics one can use steady-state data to estimate free parameters in  $G$  and  $K$ . A simple example of a chemical reaction model with a single reaction cycle is shown in Fig. 2. If we only consider the reaction scheme in the dashed box, we have the chemical species vector  $\mathbf{x} = [A \ B \ C \ D]^T$ . The stoichiometric matrix is

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

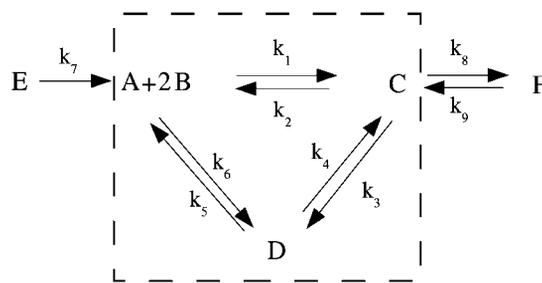


FIGURE 2 A hypothetical chemical reaction model with one cycle. The network in the dashed box is the minimal network upon which the detailed balance must be imposed.

and the rate matrix is

$$K = \begin{pmatrix} -(k_1+k_6) & k_1 & k_6 \\ k_2 & -(k_2+k_3) & k_3 \\ k_5 & k_4 & -(k_4+k_5) \end{pmatrix},$$

and  $\mathbf{g}(\mathbf{x}) = [AB^2 \ C \ D]^T$ . For a specified equilibrium vector of concentrations  $\mathbf{x}_{\text{eq}} = [A_{\text{eq}} \ B_{\text{eq}} \ C_{\text{eq}} \ D_{\text{eq}}]$ , we have

$$G = \begin{pmatrix} A_{\text{eq}}B_{\text{eq}}^2 & 0 & 0 \\ 0 & C_{\text{eq}} & 0 \\ 0 & 0 & D_{\text{eq}} \end{pmatrix}.$$

If  $K_s$  is constructed as

$$\begin{pmatrix} -(\alpha+\beta) & \alpha & \beta \\ \alpha & -(\alpha+\gamma) & \gamma \\ \beta & \gamma & -(\beta+\gamma) \end{pmatrix},$$

then by Eq. 6 the rate constant matrix that satisfies detailed balance is

$$K = \begin{pmatrix} -\frac{\alpha+\beta}{A_{\text{eq}}B_{\text{eq}}^2} & \frac{\alpha}{A_{\text{eq}}B_{\text{eq}}^2} & \frac{\beta}{A_{\text{eq}}B_{\text{eq}}^2} \\ \frac{\alpha}{C_{\text{eq}}} & -\frac{\alpha+\gamma}{C_{\text{eq}}} & \frac{\gamma}{C_{\text{eq}}} \\ \frac{\beta}{D_{\text{eq}}} & \frac{\gamma}{D_{\text{eq}}} & -\frac{\beta+\gamma}{D_{\text{eq}}} \end{pmatrix}.$$

As can be easily confirmed, this matrix satisfies the detailed balance constraint derived from the loop-based approach,  $k_1k_3k_5 = k_2k_4k_6$ .

### MINIMAL REACTION SCHEME

All chemical reactions obey detailed balance because chemistry obeys the principle of microscopic reversibility. However, for systems that are out of equilibrium, detailed

balance is violated because of nonequilibrium constraints. Many biochemical reaction systems are open systems that contain irreversible and energy-driven reactions (e.g., protein synthesis, degradation, and phosphorylation reactions, transport across a system boundary, etc.). Still, as a physical law, detailed balance must be obeyed by all reversible and non-energy-driven reactions at equilibrium. The consequence is that, in a model, detailed balance must be imposed for any non-energy-driven reaction cycle to satisfy the physical constraint even if a system may always function away from equilibrium due to constant energy dissipation or externally imposed fluxes. To impose detailed balance on the relevant part of a reaction network, it is necessary to distinguish irreversible energy-driven reactions from non-energy-driven reversible reactions. We take this information as given in the model specification.

We note that reversible reactions outside any reaction loop trivially satisfy detailed balance. Rate constants for these noncyclic reactions need not be part of an imposed detailed balance constraint. Therefore, the reaction scheme on which we need to impose detailed balance may be much smaller than the full scheme. As an example, let us consider the overall reaction scheme in Fig. 2. We can construct a binary matrix  $T$  encoding only the reaction scheme as

$$T = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The vector  $\mathbf{g}(\mathbf{x}) = [AB^2 C D E F]^T$ . To obtain the minimal reaction scheme that must satisfy detailed balance, one can remove the irreversible reaction  $E \rightarrow A + 2B$  by setting  $T_{41} = 0$  and remove the reversible acyclic reactions by setting  $T_{52} = 0$  and then  $T_{25} = 0$ . The minimal scheme is enclosed in the dashed box in Fig. 2.

Now we discuss how to automate the process of finding the minimal reaction scheme for large networks. Given a binary matrix  $T$  encoding the overall reaction scheme (i.e.,  $t_{ij} = 1$  if there is a transition from state  $i$  to  $j$  or from reaction group  $i$  to  $j$  in mass-action networks, otherwise  $t_{ij} = 0$ ) and a binary matrix  $T_e$  encoding all energy-driven reactions, we can obtain the minimal reaction scheme for imposing detailed balance by removing three types of reactions:

1. Remove all energy-driven reactions, by subtracting  $T_e$  from  $T$ , i.e.,  $T' = T - T_e$ .
2. Remove all the irreversible reactions, by setting  $T'_{ij} = 0$  if  $T'_{ij} \neq 0$  and  $T'_{ji} = 0$ . We denote the resulting matrix as  $T''$ .
3. Remove all the reversible acyclic reactions. We note that a state (or a reaction group in a mass action network) that is connected to only one other state is a state in an acyclic reaction and it can be removed from the full scheme. For any row  $i$  in  $T''$ , we set  $t''_{ij} = 0$  and  $t''_{ji} = 0$  if  $t''_{ij}$  is the only

nonzero entry in the row. Remove all such states iteratively until states in the resulting matrix  $\tilde{T}$  are connected to at least two other states.

The above procedure, however, cannot remove acyclic reactions that bridge disjoint cycles. Even though the number of such reactions is probably small in most biochemical reaction networks (e.g., it appears that there is no such acyclic reaction in the example that follows), these reactions can be removed by the following numerical method.

It is clear that acyclic reactions are always balanced at steady-state. Letting  $Q$  be an arbitrary generator consistent with the reduced scheme  $\tilde{T}$ , we can solve  $w^T Q = 0$  for the steady-state distribution  $w$ , and any reaction that satisfies  $w_i q_{ij} = w_j q_{ji}$  is an acyclic reaction and can be removed. With extremely low probability, reactions in cycles might be balanced by chance and erroneously removed because  $Q$  is arbitrary and  $w$  is solved numerically. To eliminate such cases, one can run the procedure multiple times to increase the confidence and can always verify whether detailed balance is satisfied by the full scheme. We note that  $\tilde{T}$  may contain disjoint schemes. Thus, the null space of  $Q$  may have more than one dimension. In such cases, one can construct  $w$  as a linear combination of the basis of the null space and ensure it contains no zero entries.

We have applied the above methods to find the minimal reaction network in a recently developed model for signaling by epidermal growth factor receptor (EGFR) (11). The data file and the MatLab code for this example are available online as Supplementary Material. The model contains 356 biochemical species, 1667 reaction groups, and 3749 unidirectional reactions. This model is generated by a rule-based and automated reaction network generator program called BioNetGen (12). The model has 701 energy-driven protein phosphorylations and dephosphorylations and 2460 acyclic reactions. The minimal reaction network contains 588 unidirectional reactions in cycles and 258 reaction groups, a much reduced reaction scheme. It is worth noting that the minimal network of the EGFR model contains 34 disjoint schemes (the number of the null space dimension of the minimal  $K$  matrix). In contrast to our method, a cycle-based approach would need to identify all the disjoint schemes and a fundamental cycle basis would need to be found for each one.

## CONCLUSION

Imposing detailed balance in a complex reaction mechanism using the previously proposed method (2) involves techniques to identify a cycle basis in the reaction scheme. It has been shown that finding a minimum cycle basis (which gives rise to simpler equations than does an arbitrary fundamental cycle basis) has computational complexity  $O((m+n)^{3.376})$  (13), where  $m$  and  $n$  denote the number of undirected edges (reversible transitions or reactions) and the number of vertices (states or reaction groups), respectively. For some applications with small networks, this cost may not be

significant or rate-limiting. The computational cost will be high for finding reaction cycles in a large network such as that of the EGFR model considered above. In this article, we present an alternative method of imposing detailed balance for both single molecule kinetics and nonlinear mass-action kinetics. The method provides a parameterization procedure for reaction rate constants without explicitly considering reaction cycles. We give a method to identify the minimal reaction scheme for imposing detailed balance whenever a system contains energy-driven, irreversible, and acyclic reversible reactions. For single molecule kinetics with exponential voltage dependence and mass-action ligand binding, we have shown how to fully parameterize the generator to satisfy detailed balance. This parameterization explicitly gives the correct ligand- and voltage-dependence for the equilibrium states.

## SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

We thank James R. Faeder for providing the reaction scheme data for the EGFR model.

This work was supported by National Institutes of Health grants No. GM65830 and No. RR18754 and Department of Energy contract No. W-7405-ENG-36.

## REFERENCES

1. Albery, R. A. 2004. Principle of detailed balance in kinetics. *J. Chem. Edu.* 81:1206–1209.
2. Colquhoun, D., K. A. Dowsland, M. Beato, and A. J. Pledger. 2004. How to impose microscopic reversibility in complex reaction mechanisms. *Biophys. J.* 86:3510–3518.
3. Kolmogorov, A. 1936. Theory of Markov chains. *Annal. Mathematics.* 112:155–160.
4. Fredkin, D. R., M. Montal, and J. Rice. 1985. Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer.* 1:269–289.
5. Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280.
6. Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. Lond. B Biol. Sci.* 264:375–383.
7. Bruno, W. J., J. Yang, and J. E. Pearson. 2005. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. *Proc. Natl. Acad. Sci. USA.* 102:6326–6331.
8. de Young, G., and J. Keizer. 1992. A single-pool inositol 1,4,5-trisphosphate-receptor-based model for agonist-stimulated oscillations in  $Ca^{2+}$  concentration. *Proc. Natl. Acad. Sci. USA.* 89:9895–9899.
9. Horn, F., and R. Jackson. 1972. General mass action kinetics. *Arch. Rational Mech. Anal.* 47:81–116.
10. Chaves, M., E. D. Sontag, and R. J. Dinerstein. 2004. Steady-states of receptor-ligand dynamics: a theoretical framework. *J. Theor. Biol.* 227:413–428.
11. Blinov, M. L., J. R. Faeder, B. Goldstein, and W. S. Hlavacek. 2006. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems.* 83:136–151.
12. Blinov, M. L., J. R. Faeder, B. Goldstein, and W. S. Hlavacek. 2004. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics.* 20:3289–3291.
13. Golynski, A., and J. D. Horton. 2002. A polynomial time algorithm to find the minimum cycle basis of a regular matroid. *Lect. Notes Comput. Sci.* 2368:200–209.

Jin Yang, William J. Bruno, William S. Hlavacek, and John E. Pearson

*Theoretical Biology and Biophysics Group*  
*Theoretical Division*  
*Los Alamos National Laboratory*  
*Los Alamos, New Mexico*